Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

03001637.2

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03001637.2
Demande no:

Anmeldetag:
Date of filing: 24.01.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Sony Ericsson Mobile Communications AB
Nya Vattentornet
221 88 Lund
SUEDE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Telecommunication system and method

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04Q/

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT SE SI SK TR LI

European Patent Application
SONY ERICSSON MOBILE COMMUNICATIONS AB
P27696EP

Telecommunication System and Method

The invention relates to a telecommunication system and a telecommunication method wherein a speaker's voice in converted to transmittable signals which are transmitted to a receiver. This is typical for phone connections especially with mobile and cellular phones.

When someone is talking on the phone in a noisy environment, the phone's microphone picks up not only the speaker's voice but also interfering sounds which both are converted and transmitted to the remote party, that is the receiver. The louder the interfering sounds, the more the understandability of the speaker to the remote party is reduced.

To overcome this problem, there are solutions especially with current mobile telephones wherein noise reduction algorithms are applied to the received microphone signal to improve the quality of the sound received at the remote end. Those algorithms take use of the very different frequency regions of environmental noise and of voice. However, such current noise reduction cannot cope well with environments such as noisy crowds, bars and so on, wherein the surrounding interfering noise has a very similar spectrum and loudness as the used signal, the speaker's speech, as the environment sound comprises a lot of surrounding person's voice signals.

Typically, a speaker tries to overcome this problem by more approaching to the microphone and by speaking louder. This is very often not very successful. Furthermore, video telephony applications are recently of more interest, at which applications the microphone is typically held at a viewing distance from the speaker's face and thus the microphone is farther away from the speaker's mouth. Thus, the signal to interference ratio is far more reduced.

It is an object of the present invention to enhance the signal to interference ratio even in very noisy environments.

The invention is set forth in the independent claims.

The general idea is to use a lip reading process to provide additional information within a noise reduction process. Such a lip reading process allows a feature extraction and the use thereof by detecting the lips position of the speaker only and analysing it. Based on said analysis, that is a lip reading result, the noise reduction algorithm can use additional estimates of speech energy or hints to collect statistics of the speaker's voice to better separate the speaker's voice from the surrounding sound comprising surrounding noise and surrounding persons' voices.

As in video telephony situations a camera is built in, said built-in camera can be used for detecting the lips position of the speaker and thus only the analysis and separation need additionally be done. The invention may also be used in video conference systems in which system very often more persons are sitting in one room with separate cameras and microphones, however may disturb and interfere with the microphones of the other persons.

Thus, the main advantage is that noise reduction in a phone situation can be enhanced in crowded situations.

Main and additional features of the present invention are

a) visual information, picked up by a built-in camera in the phone, is added to audio information picked up by a microphone, to form a speech signal that is transmitted to the remote party,
b) in a typical video telephony use case, the image which is already obtained in order to be transmitted to the remote party can also be used for the purpose described here,
c) real-time image processing is applied to the camera image in order to extract a small number of relevant features of the speaker's mouth,
d) the mouth features are further processed to provide input to the audio noise reduction algorithm. This could be e.g. the opening of the mouth as an estimate of speech energy; rapid movement of the mouth as a hint to consonants, etc.,
e) the noise reduction algorithm is extended to make use of features of the speaker's mouth obtained visually.

The invention is described with reference to a non-limiting typical embodiment as shown in the Fig. of the accompanying drawing.

The figure shows a typical phone apparatus 1 like video telephone having inter alia built-in a microphone M and a video camera C outputting signals which correspond

to the speech of a speaker S and the corresponding lips position of the speaker's mouth. Typically, those signals are converted to digital signals as schematically shown by A/D-converters 2 and 3, respectively, both shown by broken lines. It is to be understood, that A/D-converters 2, 3 may be incorporated into the microphone M and/or the camera C, respectively, or those may be adopted to output digital signal directly.

The signal thus corresponding to the lips position is input to an analyser 4 which is connected to a memory 5 storing typical algorithms of speech in the association to the position of the lips of a speaker. Thus the analyser 4 determines which parts of the sound, received by microphone M is coming from the speaker and which part thereof is coming from the environment, that is the environmental noise and voices of surrounding persons. The result of the analyse done by analyser 4 and the sound signal received by microphone M are input to a separator 6 which separates or distinguishes the speaker's voice from the surrounding or environmental sound and thus reduces the respective signal to interference ratio.

From the foregoing description follows that the signal corresponding to the lips position may be derived or processed both continuously and intermittently, wherein the frequency of intermittently using such signals may be variable from reached result of analysis or separation.

The output signal of the separator 6, that is a signal containing mostly only signals corresponding to the speaker's voice and having reduced if not cancelled sound from the environment, is sent to a typical converter 7 which converts this signals to transmittable signals according to various standards, the output of which converter 7 is transmitted as symbolised by a transmission 8 shown in broken lines the output signal of the camera C, a video signal, may also be input to the converter 7 and converted to transmittable signals as typical with video telephone systems or with video conference systems.

It should be noted expressively that apparatus 1 need not be a mobile or cellular video telephone apparatus and that the invention is also applicable to any system comprising a microphone and a camera.

Further, the apparatus 1 according to the invention may also comprise a learning system wherein a learning program symbolised by a block 9 communicates with the memory 5 of the algorithm of the speech. Thus, the reduction of interfering noise can be further enhanced. E.g. initially, that is before using the apparatus 1 in the

environment, the speaker S speaks given sounds, that is vocals and consonants, after having switched on apparatus 1 in a very silent environment. Thus the analyser "knows" the speaker's lip positions with predetermined consonants and vocals and thus a better separation can be reached. This learning program can also be used at the very beginning of a communication to be done.

European Patent Application
SONY ERICSSON MOBILE COMMUNICATIONS AB
P27696EP

Claims

1. Telecommunication system comprising
a microphone (M) receiving audio including a speaker's voice and environmental sound, and
a circuit (7) for converting said received audio to transmittable signals to be transmitted to a remote receiver,
wherein a noise reduction circuit reduces the environmental sound received by said microphone (M) before transmitting signals to said receiver,
wherein a detecting means (C) detects continuously or intermittently the lips position of the speaker (S),
an analysing means (4) determines speech characteristics of the speaker's voiced based on the detected lip position,
a separating means (6) separates the speaker's voice from environmental sound including environmental noise and surrounding persons' voices based on the determined speech characteristics, and
the noise reduction circuit reduces the environmental sound based on the separated environmental sound.

2. Telecommunication method wherein received audio including a speaker's voice and environmental sound is converted to transmittable signals to be transmitted to a receiver, comprising reducing environmental sound, wherein
the lips position of the speaker (S) is detected continuously or intermittently,
speech characteristics of the speaker (S) are determined by analysing the detected lips position,
the speaker's voice is separated from the environmental sound including environmental voice and surrounding persons' voices, based on said determined characteristics and
the separated speaker's voice is converted to transmittable signals.

3. The telecommunication system according to claim 1 and the telecommunication method according to claim 2, wherein the lips position of the speaker (S) is detected by means of a camera (C) viewing to the speaker's face.

4. System and method according to claim (3) wherein the camera (C) is a built-in camera of a video phone apparatus.

5. Method according to anyone of claims 2 to 4 wherein the speech characteristics are based on the opening of a speaker's mouth as an estimate of the speech energy, rapid movement of the speaker's mouth as a hint to consonants and other statistically detected characteristics of an association between position and movement of the lips of a person and the thus output voice of the speaker.

6. Method according to claim 5 wherein initially or over time of use a learning procedure is used to enhance the steps of determination and separation.

European Patent Application
SONY ERICSSON MOBILE
COMMUNICATIONS AB
P27696EP

Abstract

There is provided a telecommunication system and method wherein not only the speaker's voice but also the lips position of the speaker's mouth is detected (M; C) and is analysed (4). Based thereon the speaker's voice can be separated (6) more efficiently from environmental sound including both environmental noise and surrounding persons' voices.

(Fig. 1)

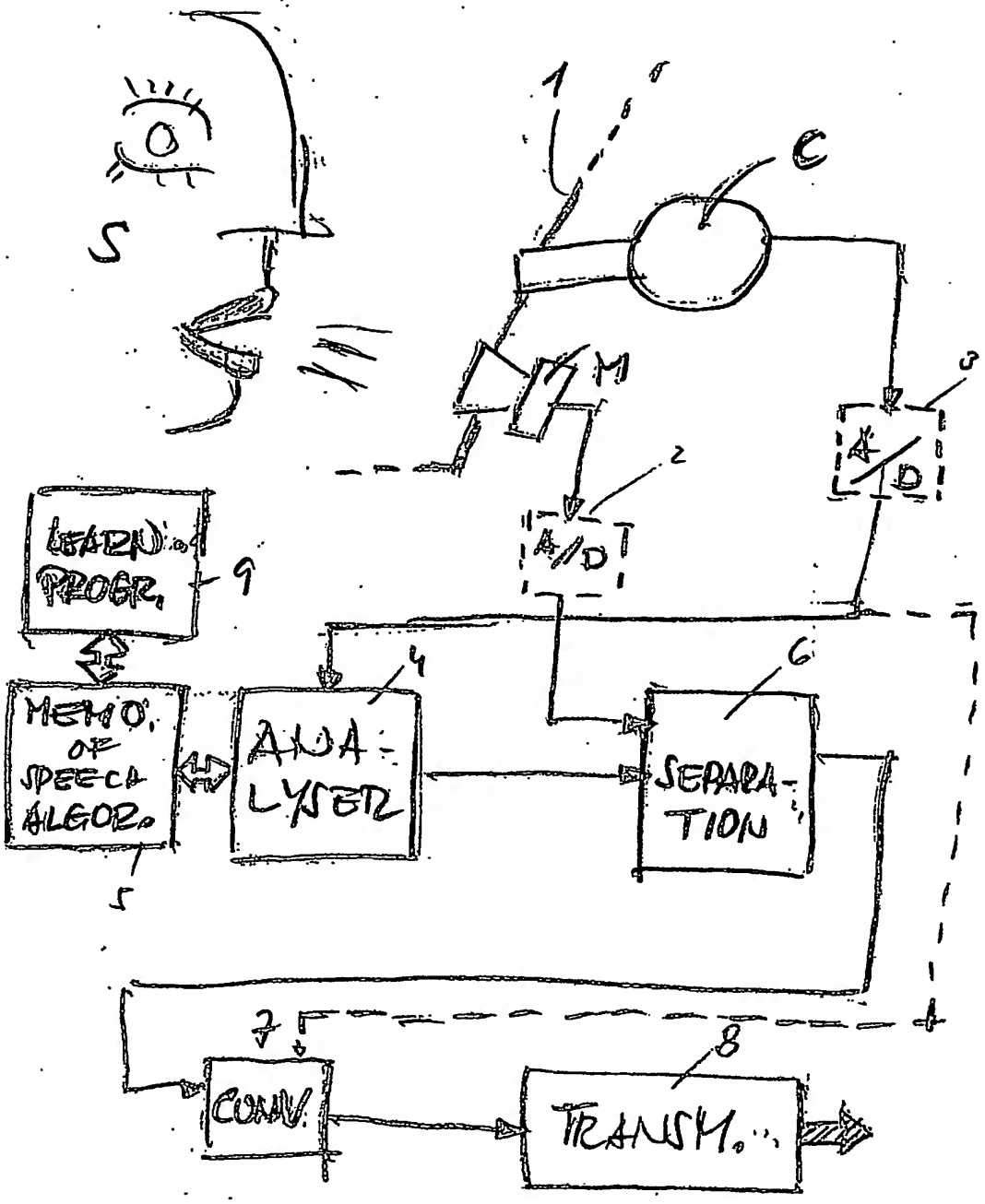

Fig. 1

This Page is inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning documents *will not* correct images problems checked, please do not report the problems to the IFW Image Problem Mailbox